DOCUMENT RESUME

ED 291 104 CS 211 061

AUTHOR Froese, Victor
TITLE Language Assesment: What We Do and What We Should Do!
PUB DATE May 87
NOTE 9p.; Paper presented at the Northwest Regional Conference of the National Council of Teachers of English (Vancouver, British Columbia, Canada, May 2, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Guides - Classroom Use - Guides (For Teachers) (052)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Elementary Secondary Education; *Evaluation Methods; Foreign Countries; Formative Evaluation; Holistic Evaluation; *Instructional Effectiveness; Language Tests; *Measurement Techniques; *Standardized Tests; Summative Evaluation; Teacher Made Tests; Teaching Methods; Theory Practice Relationship

ABSTRACT

Inconsistencies develop in the educational system both because teachers receive "mixed signals" and because theory and practice become discrepant, a situation which creates dilemmas in language assessment. Some current dilemmas include: (1) the difficulty in determining whether teacher-made and standardized tests measure what they are supposed to; (2) the fact that teachers tend to rely on product rather than process to infer students' ability; (3) the issue of whether holistic or discrete-point assessment is more beneficial; (4) the fact that teachers sometimes measure comprehension without providing comprehension instruction; and (5) the problem that standardized tests may be useful for grading but are of far less use for instruction than other measures of ability. A number of procedures which focus on the formative, on teaching, on the holistic, on the student and on process within the language arts--listening, speaking, reading and writing--are available and should be sought by the classroom teacher since ill-considered tests can lead to a thoughtless curriculum as teachers teach to inappropriate tests. (Fourteen references are attached.) (NH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

BEST COPY AVAILABLE

ED291104

LANGUAGE ASSESSMENT: WHAT WE DO AND WHAT WE SHOULD DO!

Dr. Victor Froese, Department of Language Education, UBC

Paper presented at the NCTE Northwest Regional Conference, Vancouver, British Columbia, May 2, 1987.

INTRODUCTION

Walter Loban (1976, Frontispiece) in his now classic study Language Development, Kindergarten through Grade Twelve concluded that "the language arts curriculum inevitably shrinks or expands to the boundaries of what is evaluated." Indeed, this same concern recently prompted the Committee to Improve the Teaching of English (CITE) to inventory the testing practices in British Columbia. These concerns are well founded. In fact, across Canada most Ministries of Education have separate assessment and curriculum development branches which operate independently and propose provincial assessments or provincial curricula quite out of synchronization. It is, therefore, rather easy for inconsistencies to develop in the educational system both because teachers receive "mixed signals" and because theory and practice become discrepant. And for that reason I have selected the subtitle-- What we do and what we should do! What I hope to do then is to first sketch out some of the dilemmas in language assessment and to present a few exemplary practical approaches to assessment in each of the language arts areas--listening, oral language, reading, and writing.

CURRENT DILEMMAS

Standardized vs Teacher-Made

For example, the CITE group's "Survey of the Testing of English/Language Arts in British Columbia, 1984-1985" indicates a heavy reliance on standardized tests such as the Gates-MacGinitie Reading Test and the Canadian Test of Basic Skills. Teacher-made tests are also relied on to test various objectives. The report gives an example of a comment offered by one respondant: "We rely on teacher tests that test our own objectives. This test [the standardized test] we realize gives us very limited information" (p.19). The first dilemma, then, is whether the instrument tests what it is supposed to test; that is, does it have validity. Can a test constructed on the assumption of discrete skills measure language holistically? Can such a test measure whether someone will read if not forced to read by the test situation? Does it test writing or only knowledge about writing--metalinguistics? The short answer is--if the test is not valid--don't use it. A second aspect of this dilemma is whether we want to know something about how a performance compares to a norm or whether we want some information about the individual. Standardized tests tell us little about the latter.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Victor Froese

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CS211061

Product vs Process

For decades we have used the essay as a way of testing and grading writing ability. My own children wrote one or two essays in each of their high school grades; were awarded a mark; and that was it! Occasionally they had to make an oral report or give an impromptu talk and then were given a grade. Yes, they did take standardized reading tests and we were given a grade-equivalent, if we pressed to know how they did. Listening? The reports were even more vague. The second dilemma is that we have tended to rely on products to infer ability rather than to observe the process which could tell us something about how to help the student.

Teacher vs Student-oriented Assessment

Most school work is teacher initiated; most questions are asked by the teacher; most talk in the classroom is the teachers'. Yet we know that ones "purpose" for doing things is the best motivating factor. We value "independence" yet foster dependence. The third dilemma, then, is that we must clarify the purpose for language learning. Is it to foster critical thinking, independent thought, and perhaps enjoyment? Or is it to produce language "consumers"? Each position has its appropriate methodolgy and it is incumbant on us to select the appropriate one.

Discrete-point vs Holistic Assessment

Have you ever taken a dictation test (either in your first or second language)? Have you taken a test in which you were to supply the correct inflection of a word within a sentence? These two examples illustrate holistic and discrete-point assessments respectively. The difference is, of course, that one takes language apart, the other puts it together. That is the fourth dilemma, and it is a troublesome one. We all have seen the results of teaching foreign languages and composition through grammar. That is why we use "immersion" approaches today for the former and "process writing" for the latter. Oller (1979, p.60) suggests that the argument for discrete-point testing has been weakened by the fact that integrative (i.e.holistic) tests correlate better with each other even though they may be as diverse as a cloze test or a dictation test.

Testing vs Teaching

Often when teaching a university class in Reading, I make the statement that I have never seen a comprehension lesson being taught. It is close to the truth and it is corroborated by research evidence. Durkin (1978-79,pp.481-533) concluded after observing Social Studies classes that in Grades 3 through 6 almost no comprehension instruction was found. At first such

statements appear shocking but a little reflection will indicate that we spend very little time modelling and considerable time in testing. For example, we assign an essay and provide a grade; we ask questions about a novel and correct them; we give directions about assignments and expect results. I have rarely seen a teacher model the process of arriving at the result--a process aptly named "scaffolding" by Applebee & Langer (1984). The contrast between testing and teaching is dilemma number five.

Summative vs Formative Assessment

We use some forms of assessment for the purpose of deriving grades and some for modifying instruction. Under some circumstances this might lead to a "chicken-and-egg" argument but the point I wish to make is that information obtained for different purposes often is different in nature. While standardized test results may be useful for grading, ~~it is~~ they are of far less use for instruction. What does a percentile, stanine, or grade-equivalent score in vocabulary or spelling tell us that is directly useful for instruction? On the other hand , knowing that a student cannot correctly punctuate a sentence containing a series of events, is of some help. The sixth dilemma is, therefore, that the purpose for testing may influence the form of testing used.

CLASSROOM APPLICATIONS

Within the framework of the above concepts, I will now outline a number of procedures which focus on the formative, on teaching, on the holistic, on the student, and on process within the language arts--listening, speaking, reading, and writing.

Listening Assessment

1. The Wilkinson Framework

Wilkinson (1974, p.66) presented a useful framework which may be used for studying almost any piece of communication: " Who communicates what to whom, how and why, and on what occasion?" These six questions can make a very useful informal observation form for the teacher.

2. Dictation & Transcription

Moffett (1983, p.490) suggests that "transcribing is writing without composing" and recommends taking dictations from classmates, acting as a scribe for a group, taping own ideas and then transcribing, and taking down live speech (very difficult).

3. Retelling

Ideally a passage should be read by one person to the audience and retold to a different person so as to retain some validity in

the task. A checklist of key concepts may be used to assess the retelling.

Speaking & Oral Language Assessment

4. Grouptalk

For the first several times, divide the class into small groups (5-7 students) and set up a taperecorder for each group to be left on during the session. The tape is played back before the "evaluation" is done. It may also be necessary to insist that every student contribute before the responding begins. The discussion, especially when led by a student, might begin with a question written by the teacher on an index card or on chart paper so everyone can see it.

5. Directed Reading-Thinking Activity

The students need an essay, a story, or a content area piece to use as described below.

a. The teacher needs to analyze the piece for natural breaks where discussion should occur.

b. A student is appointed as recorder for each group.

c. Students examine the title and make predictions about what will follow. These predictions are written down by the recorder. A similar procedure is followed at each stop.

d. Students read the text to confirm or disprove their predictions. These are then discussed at each stop.

e. The teacher's role is to keep the discussions moving by asking general, non-cueing questions such as: What do you think will happen? Why do you think so?

6. Small Group Interview

Using about three students, each reads a passage presented (30%), then they engage in a discussion (40%), and in a conversation with the examiner (30%). (Based on London University Spoken English Test described in Knowles (1983))

7. Inner-Outer Circle

A small inner circle discusses a given topic while the outer circle observes and evaluates discussion (using an appropriate checklist).

Checklist might include:
(Based on Knowles, 1983, p.236)

a. interest aroused in subject
b. use of vocabulary and idiom
c. quality of replies to questions
d. ability to listen to others
e. ability to communicate thoughts and feelings from written word.

Reading process assessment

8. Oral Reading & Miscue Analysis

The following suggestions are adapted from Moffett (1983, p. 461):

a. Which elements of the text are ignored (phonemes, word endings, whole words, phrases, punctuation)?

b. Which spellings are sounded out incorrectly (spelling patterns not systematic or standard)?

c. Which combinations of sound-spellings trip up the reader (blends, vowel-consonant combinations, polysyllabic words)?

d. Are punctuation, capitalization, or headings followed?

e. Which kinds of mistakes are corrected? Are they important to meaning?

f. Are sounds, words, or other elements rearranged?

g. Which substitutions seem to constitute "reading into the text" subjective expectancies, preoccupations, or stereotypes?

h. Whether phrasing and intonation fit the sense as well as the syntax and punctuation?

i. Is the reader "involved" or is the reading mechanical?

9. Cloze Tests

A variety of cloze tests, depending on age and purpose may be constructed to assess the instructional level of the student or some more specifc aspects of text such as anaphoric relations, concepts, or transitional relationships.

To construct a cloze test delete every fifth (or seventh, or tenth) word and replace with a standard length blank. For primary grades alternatives for each blank may be listed; for intermediate and above that may not be necessary. Aulls (1982) provides examples for a number of cloze variations: ZIP cloze, content & synonym cloze, maze cloze, structure word cloze, content word cloze.

10. Scaffolding

Five aspects of natural language learning are emphasized: intentionality, appropriateness, structure, collaboration, and internalization.

Assessment of Writing & Composing

11. Writing Folder

Graves (1983) suggests that a writing folder is the only feasible method of record keeping since any procedure must be useful for the teacher and the student. He also describes the changing nature of the method of record keeping (Graves, 1983, p.307-308). The folder could initially contain information on all four sides: Side 1 would contain a record of dates when each piece of writing was begun and was completed; Side 2 would contain topics of interest to the writer; Side 3 would contain a record of mastered writing skills; and Side 4 would list the writer's areas of writing expertise. As needed the teacher would add other types of records such as details of conferences, students' responses to questions during the sharing times, class progress charts, and results of pretests based on dictation exercises. This is not a static form of record keeping but instead one that is adapted to the needs of teacher and student--some aspects are relatively permanent, others rather temporary. Since the purpose of record keeping and assessment is to assist the student in the "process" of writing it must be meaningful to be retained. Some procedures outlive their usefulnes and hence are discontinued.

12. Holistic Scoring

This method of scoring compositions may accomplish one of several purposes described by Cooper (1977, p.3):

a. It may be matched to another piece of writing in a graded series,

b. It may be scored for the prominence of certain features important to that kind of writing,

c. It may be assigned a letter or number grade.

The grading is quick (i.e. two minutes) and impressionistic and is intended to provide general feedback to the writer. If features are involved each is rated high, middle, or low. Cooper (1977) defends this procedure as following: "A piece of writing communicates a whole message with a particular tone to a known audience for some purpose: information, argument, amusement, ridicule, titilation. At present, holistic evaluation by a human respondent gets us closer to what is essential in such a

communication . . . (p. 3)"

Details of holistic evaluation are found in Cooper's (1977) "Holistic evaluation of writing," in Myers'(1980) booklet A Prodedure for Writing Assement and Holistic Scoring, and in White's (1985) Teaching and Assessing Writing. Three types of evaluation are commonly used:

a. General Impression Marking - The rater simply scores the paper by deciding where the paper fits within the range of papers produced for that assignment or occasion.

b. Primary Trait Scoring - Focusses on just those features of a piece which are relevant to the kind of discourse it is.

c. Analytic Scoring - Each feature to be evaluated is described in some detail together with high-mid-low points, e.g. ideas, organization, wording, etc.

13. Writing Exercises

See Duke, C.R. (Ed).(1984). Writing exercises from Exercise Exchange. Vol. II. Urbana, Illinois: NCTE.

This is a junior high to college level collection of sources for writing, prewriting, modes for writing, writing & reading, language mechanics and style, and revising/responding/evaluating.

CONCLUSIONS

While the foregoing is only a small set of the procedures available to the teacher today, it is suggested that is a set of ideas compatible with the notions highlighted through the six dilemmas. As White (1985, p.8-9) has stated: "A theory of knowledge of the subject must precede any practice of testing that is to make coherent sense and that is to be useful to those teaching and learning: indeed, when theories of knowledge and content are not taken into consideration, ill-considered tests can lead to a thoughtless curriculum as teachers wind up teaching to inappropriate tests."

BIBLIOGRAPHY

Applebee, A.N. and Langer, J.A. (1984). Instructional scaffolding: Reading and writing as natural language activities. In Jensen, J.M. (Ed.). Composing and Comprehending. Urgana, Illinois, ERIC & NCRE, pp.185-190.

Aulls, M. (1982). Developing Readers in Today's Elementary School. Toronto: Allyn & Bacon, pp.355-363.

Committee to Improve the Teaching of English. (1986). Survey of the Testing of English/Language Arts in British Columbia, 1984-1985.

Cooper, R.L. and Odell, L. (1977). Evaluating Writing. Urbana, Illinois: NCTE, p.3.

Duke, C.R (Ed.).(1984). Writing Exercises from Exercise Exchange. Vol. II. Urbana, Illinois : NCTE.

Durkin, D. (1978-79). What classroom observations reveal about reading comprehension instruction. Reading Research Quarterly, 14, 4, pp.481-533.

Graves, D.H. (1983). Writing: Teachers and Children at Work. Exeter, New Hampshire: Heineman.

Knowles, L. (1983). Encouraging Talk. London: Methuen.

Loban, W. (1976). Language Development: Kindergarten Through Grade Twelve. Urbana, Illinois: National Council of Teachers of English, frontispiece.

Moffett, J. and Wagner, B.J. (1983). Student-Centered Language Arts and Reading, K-13: A Handbook for Teachers. Boston: Houghton Mifflin.

Myers, M. (1980). A Procedure for Writing Assessment and Holistic Scoring. Urbana, Illinois: ERIC & NCTE.

Oller, J.W. (1979). Language Tests at School. London: Lognman, p.60.

White, E.M. (1985). Teaching and Assessing Writing. San Francisco: Jossey-Bass Publishers.

Wilkinson, A., Stratta, L. and Dudley. P. (1974). The Quality of Listening. London: Macmillan.